Presmoothed Landmark estimators of the transition probabilities

Luís Meira-Machado[1]

[1] Centre of Mathematics & Department of Mathematics and Applications, University of Minho, Campus de Azurem, 4800-058 Guimarães, Portugal.

E-mail for correspondence: lmachado@math.uminho.pt

Abstract: Multi-state models can be successfully used to model complicated event history data, for example, describing stages in the disease progression of a patient. In these models one important goal is the estimation of the transition probabilities since they allow for long term prediction of the process. There have been several recent contributions for the estimation of the transition probabilities. Recently, de Uña-Álvarez and Meira-Machado (2015) proposed new estimators for these quantities, and their superiority with respect to the competing estimators has been proved in situations in which the Markov condition is violated. In this paper, we propose a modification of the estimator proposed by de Uña-Álvarez and Meira-Machado based on presmoothing. Simulations show that the presmoothed estimators may be much more efficient than the completely nonparametric estimator.

Keywords: Kaplan-Meier; Multi-state model; Nonparametric estimation; Transition probabilities.

1 Introduction

In many medical studies individuals can experience several events across a follow-up study. Analysis of such studies can be successfully performed using a multi-state model (Meira-Machado et al., 2009). This paper introduces and studies a feasible estimation method for the transition probabilities in a progressive multi-state model.

Fully non-Markov estimators for the transition probabilities were introduced for the first time in Meira-Machado et al. (2006). Recently, this problem has been reviewed, and new sets of estimators have been proposed (de Uña-Álvarez and Meira-Machado, 2015). This method proceeds by considering specific subsets of individuals (namely, those observed to

This paper was published as a part of the proceedings of the 31st International Workshop on Statistical Modelling, INSA Rennes, 4–8 July 2016.

be in a given state at a pre-specified time point) for which the ordinary Kaplan-Meier survival function leads to a consistent estimator of the target. Superiority with respect to the competing estimators has been proved.
A multi-state model is a stochastic process $(X(t), t \in \mathcal{T})$ with a finite state space, where $X(t)$ represents the state occupied by the process at time t. For simplicity, in this paper we assume the progressive illness-death model and we assume that all the subjects are in State 1 at time $t = 0$. The illness-death model describes the dynamics of healthy subjects (State 1) who may move to an intermediate "diseased" state (State 2) before entering into a terminal absorbing state (State 3). Many longitudinal medical data with multiple endpoints can be reduced to this structure.
The illness-death model is characterized by the joint distribution of (Z, T), where Z is the sojourn time in the initial state 1 and T is the total survival time. Both Z and T are observed subject to a random univariate censoring C assumed to be independent of (Z, T). Due to censoring, rather than (Z, T) we observe $(\widetilde{Z}, \widetilde{T}, \Delta_1, \Delta_2)$ where $\widetilde{Z} = \min(Z, C)$, $\Delta_1 = I(Z \leq C)$, $\widetilde{T} = \min(T, C)$, $\Delta_2 = I(T \leq C)$, where $I(\cdot)$ is the indicator function. The target is each of the five transition probabilities $p_{ij}(s,t) = P(X(t) = j \mid X(s) = i)$, where $1 \leq i \leq j \leq 3$ and $s < t$ are two pre-specified time points.

2 Estimators

The transition probabilities are functions involving expectations of particular transformations of the pair (Z, T). In practice, we only need to estimate three transition probabilities since the others can be expressed from these ones; namely,

$$p_{11}(s,t) = \frac{E\left[I(Z > t)\right]}{E\left[I(Z > s)\right]}, \qquad p_{13}(s,t) = \frac{E\left[I(s < Z, T \leq t)\right]}{E\left[I(Z > s)\right]},$$
$$p_{23}(s,t) = \frac{E\left[I(Z \leq s < T \leq t)\right]}{E\left[I(Z \leq s < T)\right]}.$$

Because of space limitation we will focus on the transition probability $p_{13}(s,t) = P(T \leq t \mid Z > s)$. Given the time point s, to estimate this quantity, the analysis can be restricted to the individuals with an observed first event time greater than s. This is known as the landmark approach (van Houwelingen et al. 2007). The corresponding estimator (KMW) is given by $\widehat{p}_{13}(s,t) = 1 - \widehat{S}_T^{(s)}(t)$ where $\widehat{S}_T^{(s)}(t)$ denote the Kaplan-Meier estimator computed from the given sub sample. Similarly, the transition probability $p_{23}(s,t) = P(T \leq t \mid Z \leq s < T)$ can be estimated by considering specific subsets of individuals (namely, those observed to be in a state 2 at a pre-specified time point s, i.e. those for which $Z \leq s < T$) for which the ordinary Kaplan-Meier survival function leads to a consistent estimator (see de Uña-Álvarez and Meira-Machado (2015) for further details).

The standard error of the estimators introduced by de Uña-Álvarez and Meira-Machado may be large when censoring is heavy, particularly with a small sample size. Interestingly, the variance of this estimator may be reduced by presmoothing. This 'presmoothing' is obtained by replacing the censoring indicator variables in the expression of the Kaplan-Meier weights (used by the authors), by a smooth fit. This preliminary smoothing may be based on a certain parametric family such as the logistic, or on a nonparametric estimator of the binary regression curve. The (semiparametric) Kaplan-Meier Presmoothed Weighted estimator (KMPW) is given by $\widehat{p}_{13}^{\star}(s,t) = 1 - \widehat{S}_T^{(s\star)}(t)$ where $\widehat{S}_T^{(s\star)}(t)$ denotes the presmoothed Kaplan-Meier estimator in the same sub sample.
Note that, unlike the estimator by de Uña-Álvarez and Meira-Machado (2015), the semiparametric (presmoothed) estimator can attach positive mass to pair of event times with censored total time. However, both estimators attach a zero weight to pairs of event times for which the first event time is censored. In the limit case of no presmoothing, the two estimators are equivalent.

3 Simulation study

In this section we investigate the performance of the proposed estimators through simulations. The simulated scenario is the same as that described in Amorim et al. (2011). To compare the performance of the methods we compute the mean square error (MSE), bias and standard deviation (SD). For completeness we also included the estimator by Meira-Machado et al. (2006).
Figure 1 depicts the boxplots of the estimated MSE over 1000 simulated datasets. From this plot it can be seen that with exception of the (LIDA) estimator by Meira-Machado et al. (2006) the remaining two estimators (KMW and KMPW) perform well, approaching their targets as the sample size increases. Besides, simulation results also reveal that the new proposal perform favorably when compared with the competing methods. Our simulation results reveal relative benefits of presmoothing in the heavily censored scenarios or small sample sizes.

Acknowledgments: This project was funded by FEDER Funds through "Programa Operacional Factores de Competitividade - COMPETE" and by Portuguese Funds through FCT - "Fundação para a Ciência e a Tecnologia", in the form of grant PEst-OE/MAT/UI0013/2014.

References

Amorim, A. P., de Uña-Álvarez, J., and Meira-Machado, L. (2011). Presmoothing the transition probabilities in the illness-death model. *Statis-*

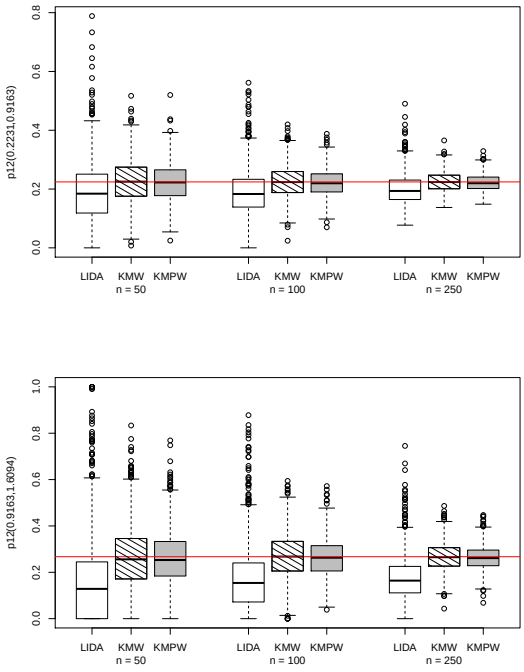

FIGURE 1. Mean square error for the three estimators.

tics & Probability Letters, **81**, 797 – 806.

de Uña-Álvarez, J., and Meira-Machado, L. (2015). Nonparametric estimation of transition probabilities in the non-Markov illness-death model: a comparative study *Biometrics*, **71(2)**, 364 – 375.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457 – 481.

Meira-Machado, L., de Uña-Álvarez, J., and Cadarso-Suárez, C. (2006). Nonparametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Analysis*, **12**, 325 – 344.

Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C., and Andersen, P.K. (2009). Multi-state models for the analysis of time to event data *Statistical Methods in Medical Research*, **18**, 195 – 222.

van Houwelingen, H.C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, **34**, 70 – 85.